# How do we reduce clinically irrelevant alarms in intensive care units? A prospective, observational, clinical study

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID: | bmjopen-2013-003354 |
| Article Type: | Research |
| Date Submitted by the Author: | 07-Jun-2013 |
| Complete List of Authors: | Inokuchi, Ryota; The University of Tokyo Hospital, Emergency and Critical Care Medicine<br>Sato, Hajime; National Institute of Public Health, Health Policy and Technology Assessment<br>Nanjo, Yuko; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Echigo, Masahiro; Joint Graduate School of Tokyo Women's Medical University and Waseda University, Cooperative Major in Advanced Biomedical Sciences<br>Tanaka, Aoi; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Ishii, Takeshi; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Matsubara, Takehiro; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Doi, Kent; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Gunshin, Masataka; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Hiruma, Takahiro; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Nakamura, Kensuke; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Shinohara, Kazuaki; Ohta Nishinouchi Hospital, Emergency and Critical care medicine<br>Kitsuta, Yoichi; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Nakajima, Susumu; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Umezu, Mituso; Joint Graduate School of Tokyo Women's Medical University and Waseda University, Cooperative Major in Advanced Biomedical Sciences<br>Yahagi, Naoki; The University of Tokyo Hospital, Emergency and Critical care medicine |
| Primary Subject Heading: | Intensive care |
| Secondary Subject Heading: | Health informatics, Nursing |
| Keywords: | Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Adult intensive & critical care < ANAESTHETICS, Health |

| | |
|---|---|
| | informatics < BIOTECHNOLOGY & BIOINFORMATICS |
| | |

**How do we reduce clinically irrelevant alarms in intensive care units? A prospective, observational, clinical study**

**Running Title**: Reduce clinically irrelant alarms in ICU

Ryota Inokuchi MD[1], Hajime Sato MD, MPH, DrPH, PhD[2*], Yuko Nanjo MHS, RN[1], Masahiro Echigo[3] , Aoi Tanaka RN[1], Takeshi Ishii MD[1], Takehiro Matsubara MD, PhD[1], Kent Doi MD, PhD[1], Masataka Gunshin MD[1], Takahiro Hiruma MD[1], Kensuke Nakamura MD[1], Kazuaki Shinohara MD, PhD[4], Yoichi Kitsuta MD, PhD[1], Susumu Nakajima MD, PhD[1], Mitsuo Umezu PhD[3], and Naoki Yahagi MD, PhD[1]

[1] Department of Emergency and Critical Care Medicine, The University of Tokyo Hospital 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

[2] Department of Health Policy and Technology Assessment, National Institute of Public Health 2-3-6 Minami, Wako, Saitama 351-0197, Japan

[3] Cooperative Major in Advanced Biomedical Sciences, Joint Graduate School of Tokyo Women's Medical University and Waseda University, TWIns, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan

[4] Department of Emergency and Critical Care Medicine, Ohta Nishinouchi Hospital, 2-5-20 Nishinouchi, Koriyama, Fukushima 963-8558, Japan

**Grant:** This work was supported by Grant-in-Aid for Young Scientists (C) (127100000424), and a Health Labour Sciences Research Grant.

**Acknowledgments**

The authors are deeply grateful to Yugo Tamura for collecting data, and would like to thank Yohei Hashimoto, Kikuo Furuta, and Hiroko Hagiwara for their support. The authors would also like to thank all participating ICU members at The University of Tokyo Hospital for their support.

**Address correspondence to:** Hajime Sato, MD, MPH, DrPH, PhD: Department of Health Policy and Technology Assessment, National Institute of Public Health, 2-3-6 Minami, Wako, Saitama 351-0197, Japan: E-mail: sato.inokuchi2@gmail.com

**Summary**

*Article Focus*

Our objective was to determine whether or not patient clinical severity affects the proportion of clinically relevant alarms and if the proportion and number of clinically relevant alarms differs based on the type of the monitoring device.

*Key Messages*

・In this study we demonstrated that clinically relevant alarms decrease proportionally as the clinical severity decreases

・The monitoring devices that triggered alarms the most often were the ART (38.0%), ECG (25.7%), and $SpO_2$ monitors (21.6%). Unfortunately, they generated less than 20% of the relevant alarms. For instance, only 6% of the alarms triggered by the ECG were clinically relevant, and 77.9% of the alarms were induced or false.

・Clinically irrelevant alarms were reduced by 28% by evaluating their technical relevance.

*Strengths and Limitations*

Prior to this study record data from a 24-h video monitor, with the help of 2 physicians, to evaluate the clinical relevance of the alarms. This technique reduced the possible bias introduced by bedside evaluation. The same method of evaluation was used in this study, with the added evaluation of the fluctuations in alarm relevance and clinical severity in individual patients. The present study also extends the previous work to a larger cohort. Although a determination could be made regarding whether an alarm was *technically true* or *false*, a strict definition of clinical relevance was more difficult. There are clinically relevant alarms that require an immediate response and alerting alarms that are meaningful but do not require an immediate response. Inter-rater agreement also varied regarding these alarms.

*STUDY LIMITATIONS*

This study has several limitations. The first is our small sample size of 13 patients. The second limitation is that although a determination could be made regarding whether an alarm was *technically*

*true* or *false*, a strict definition of clinical relevance was more difficult. There are clinically relevant alarms that require an immediate response and alerting alarms that are meaningful but do not require an immediate response. Inter-rater agreement also varied regarding these alarms.

**Abstract**

**Background:** Irrelevant alarms in an intensive care unit (ICU) can harm patient safety. Here, we determined whether the proportion and number of clinically relevant alarms differs based on the type of monitoring device used and patient illness severity, and to suggest a method for reducing clinically irrelevant alarms.

**Methods:** We conducted a prospective, observational, clinical study in a medical ICU of a university hospital. Between February and March 2012, 13 ICU patients were monitored. The alarms, alarm settings, alarm messages, waveforms, and video recordings were acquired in real-time and saved continuously. The alarms were annotated with respect to technical validity and clinical validity.

**Results:** During 2352 person-hours of continuous monitoring, 8013 alarms were annotated. Approximately 27.6% of these alarms were classified as technically false. Only 534 (6.7%) alarms were considered clinically relevant. Most of the generated alarms were threshold alarms (70%). Direct measurements of arterial pressure (38.0%), electrocardiograms (25.7%), and oxygen saturation (21.6%) measurements triggered most alarms. Positive correlations were established between patient clinical severities and the proportion of relevant alarms. The total number of irrelevant alarms could be reduced by 28% by evaluating their technical relevance.

**Conclusions:** We demonstrated that the proportion of clinically relevant alarms decreased as the patients' status improved, and the total number of alarms can be considerably reduced by evaluating their technical relevance.

**Keywords:** monitoring alarms; intensive care unit; alarm algorithm; clinical decision support; patient safety

**Background**

In the intensive care unit (ICU) setting, a large number of medical devices are attached to patients, generating numerous alarm signals every day. Several studies have demonstrated that most of these alarms are not clinically relevant [1-3] and tend to lower the attentiveness of the medical staff, and in turn, lower patient safety [4,5]. In addition, alarm sounds are associated not only with patient delirium [6-10], which increases mortality [11], but also with medical staff memory and judgment disturbances, decreased sensitivity, and exhaustion [6,7]. Many attempts have been made to reduce the number of clinically meaningless alarms through the use of statistical methods and artificial intelligence systems [12]. Some examples include extending the time period between the incident and the sounding of the alarm, shutting off alarms prior to performing procedures on patients, combining multiple clinical parameters to reduce the number of separate alarms, and calibrating machines to detect gradual changes in the patient condition. However, discrepancies remain between the priorities of equipment manufacturers who are seeking devices with high sensitivity and those of medical professionals who desire machines with high specificity.

Previous studies have demonstrated that of the 3 types of alarms—threshold alarms, arrhythmia alarms, and technical alarms—clinical relevance is the lowest for threshold alarms [13]. Our objective was to determine whether patient clinical severity affects the proportion of clinically relevant alarms and if the proportion and number of clinically relevant alarms differ based on the type of the monitoring device. To answer these questions, we used video monitors to collect 24-h continuous data from ICU patients.

**Materials and Methods**

*STUDY SETTING AND PATIENT POPULATION*

This study was conducted in a 6-bed, mixed ICU at the University of Tokyo Hospital, where patients are mainly admitted following ambulance transport. The study ICU is organized in an “I” shape, with 2 individual patient rooms on the west side and 2 double patient rooms on the east side, with a central monitoring station. The doors to the patient rooms were normally left open unless procedures were being performed or privacy was required. The unit was staffed with 1 nurse for every 2 patients. Most

patients had sepsis, respiratory failure, acute respiratory distress syndrome, multisystem organ failure, renal failure, heart failure, or trauma.

The following inclusion criteria were used to enroll patients in the study: 1) directly admitted to the University of Tokyo Hospital mixed ICU, not stepped-down from other ICUs, and 2) age ≥18 years. Patients were excluded if they were 1) already admitted to this ICU, 2) admitted for <2 days, or 3) the patient refused active treatment. This study was approved by the Ethics Committee of the University of Tokyo Hospital, and all patients or their family signed an informed consent before the beginning of the recordings.

*DATA COLLECTION*

General patient information such as age, gender, and disease was recorded. All patients were continuously videotaped using a network of cameras (JVCKENWOOD, V.NET@Web, Tokyo, Japan), attached to the ceiling above each bed, to record patient and/or system manipulations. Each patient was monitored for heart rate, invasive or closely monitored noninvasive arterial blood pressure, respiratory rate, oxygen saturation ($SpO_2$), and temperature. In addition, any changes in the equipment used to monitor each patient were monitored throughout the recording period, including which alarm was active. The alarmed equipment included infusion pumps, arterial pressure monitors, $SpO_2$ monitors, electrocardiograms (ECGs), nasogastric tubes, and urinary catheters. Ventilator alarms were not considered in this study because the alarm message in ventilator was not recorded on our system. In addition, the acute physiology and chronic health evaluation (APACHE II) score [14] was calculated for each patient within 24 h of admission, and the sequential organ failure assessment (SOFA) score [15] was calculated every 8 h. Patient data were pseudonymized, and the electronic files and videos were stored in locked, encrypted, hard drives.

*ALARM SYSTEMS AND SETTINGS*

During the study period, all patients were monitored with a standard cardiovascular monitoring system (BSM-9101 & CNS-9701, Nihon Koden, Tokyo, Japan). The numerical measurements, waveforms, alarms, alarm settings, and alarm messages, were acquired in real-time and saved

continuously (CNS-9600 & CAP-2100, Nihon Koden). The alarm information consisted of the alarm grade, the parameter causing the alarm, and the alarm message. The alarm messages were divided into 3 types: threshold alarms, arrhythmia alarms, and technical alarms. The technical alarms indicated technical problems, such as a disconnected probe. The technical alarms were helpful because it indicated that a fault in the connection of the medical device to the patient, but did not require clinical judgment, and were excluded from further analysis.

The initial alarm limits and every modification of these during the observation period were registered with corresponding time stamps and automatically recorded (CNS-9600 & CAP-2100, Nihon Koden). Chambrin et al. determined the initial limits for heart rate and systolic arterial pressure by using the rule, "initial value observed during a stable period ± 30%". [1]. This rule was used in this study as well; however, when prehospital patient heart rates and arterial pressures were not obtained, initial limits were 156/56 mmHg (120/80 ± 30%) for systolic arterial pressure/diastolic pressure and 78 beats/min for upper limit, and 43 beats/min (60 ± 30%) for lower limits for heart rate, if there were no specific alarm settings. In addition, the $SpO_2$ limit was 95%, except for patients with chronic obstructive pulmonary disease or acute respiratory distress syndrome patients, where the limit was 90%; a temperature limit of 38.3°C was also used. After these initial settings, the alarm limits could be modified; any changes were automatically recorded.

Technical and Clinical Annotations

After completion of the data collection for a particular patient, 2 nurses and 2 intensivists, with at least 6 years' experience in intensive care medicine, annotated the data. The 2 nurses first analyzed the technical validity of the alarms, with the threshold and arrhythmia alarms divided into 3 categories, *technically true*, *technically false*, and indeterminable, referring to the recorded wave shapes from the monitoring and the video record. The classifications were defined according to following criteria:

Electrocardiogram, Oxygen saturation, Direct measurement of arterial pressure, and End-tidal carbon dioxide

If the waveform was obviously an artifact produced by movements or procedures, the alarm was

determined to be *technically false*. For waveforms in which the origin of the artifact(s) or arrhythmia(s) were uncertain, other waveforms or pulse rate (e.g., a direct measurement of arterial pressure (ART) or $SpO_2$) at the time of alarm generation were also referenced. Alarms that did not meet any of the above criteria were considered *technically true*. All technical evaluations that could not be determined from the relevant monitor's waveform were defined as indeterminable.

Temperature

All upper and lower limits of the temperature alarms were defined as t*echnically true*.

Noninvasive blood pressure

When apparently abnormal values were obtained for the noninvasive blood pressure (NIBP) measurements and these were caused by patient movements, which triggered upper and lower limit alarms, these alarms were considered *technically false*.

After the technical analyses, the 2 physicians divided the alarms into 3 types, a relevant alarm; a helpful, but not relevant, alarm; and irrelevant alarms. In this study, a situation was defined as alarm-relevant when an immediate diagnostic or therapeutic decision was necessary. When the situation did not require an immediate action, but the alarm was judged to be helpful, the alarm was classified as a helpful, but irrelevant, alarm (e.g., a patient with a brain hemorrhage or subarachnoid hemorrhage who demonstrated a transiently increased blood pressure). Intensivists determining the clinical relevance could see the result of technical validity, but could not see the duration of the sounding of the alarm.

*STATISTICAL ANALYSES*

All included patient characteristics were described using means and SDs for continuous variables, and medians and interquartile ranges (IQR) for original values; categorical data were reported as numbers and proportions. After obtaining the descriptive statistics regarding alarm counts and their proportions, the bivariate relationship of the alarms (the total number of alarms, and the proportions of relevant alarms/relevant plus alerting alarms) with the patient severity (SOFA) scores were examined by fitting

cross-sectional time-series models for panel data. Alarms from different monitoring devices were examined separately and together. In a preliminary analysis, the numbers and proportions of alarm types were regressed against SOFA scores by fitting both fixed-effect and random-effect models. Both models provided similar results, but the Hausman test indicated that the fixed-effect estimators were consistently more appropriate than the random-effect estimators. Therefore, the results obtained by the fixed-effect model were adopted.

The intraobserver and interobserver variability between the 2 physicians in the clinical annotations of alarms, and between the 2 nurses in the technical annotations of the alarms was judged by a Kappa test [16]. To evaluaten the intraobserver variability, 300 alarm situations were annotated again by the same observer after a period of approximately 6 months. Statistical analyses were conducted using STATA Special Edition version 12.1 (StataCorp, College Station, TX, USA).

**Results**

PATIENT CHARACTERISTICS

Between January and February 2012, a total of 15,229 alarms were recorded for 20 patients. Two patients were excluded because of their poor clinical condition at the time of admission, and their families' lack of expected benefit from invasive treatment. Five additional patients were excluded because they demonstrated improved conditions and were discharged from the ICU within 2 days. Therefore, 13 patient cases were included in this study, corresponding to 2352 person-hours of continuous monitoring. The observation time for the cases averaged 181 ± 111 h. Table 1 describes each patient's characteristics at the time of their admission. The mean age, APACHE II score, and SOFA score upon admission to the ICU were 67.9 ± 16.3 years, 17.2 ± 8.26, and 5.92 ± 4.03, respectively. During their treatment in the ICU, 69.2% of the patients improved (SOFA scores decreased), while 15.4% deteriorated (SOFA scores increased).

ALARM CLASSIFICATIONS

A total of 8013 alarms were included in the analysis, classified as *technically true* (61%), *technically false* (28%), and indeterminable (11%) alarms (Fig. 1, Table 2). The interobserver variabilities in the

technical and clinical annotations, as estimated by the κ coefficient, were 0.98 and 0.68. In the same way, the intraobserver validities were 0.95 and 0.73. These values are within the range of substantial (0.61–0.80) or almost perfect agreement (0.81–1.00). The overall contribution of each alarm type to the 8013 alarms is shown in Table 2. Only 6.7% of all alarms were actually relevant to patient care, whereas 18.2% were helpful, but not relevant, and 75.1% of all alarms were not alarm relevant. Alarm relevant of all alarms comprised *technically true* and indeterminable alarms, not those that were *technically false*. Over an 8-hour shift, a nurse would hear 1–2 relevant alarms out of a total of approximately 25 alarms.

The monitoring devices that triggered alarms the most often were the ART (38.0%), ECG (25.7%), and $SpO_2$ monitors (21.6%) (Figure 2). Unfortunately, they generated less than 20% of the relevant alarms. For instance, only 6% of the alarms triggered by the ECG were clinically relevant, and 77.9% of the alarms were induced or false.

EFFECT OF PATIENT STATUS ON THE ALARMS

The results of the cross-sectional time-series analysis are shown in Table 3. SOFA scores had statistically significant positive coefficients when regressed against the total number of alarms, as well as against the proportion of relevant and relevant plus helpful alarms ($p < 0.05$). They indicated that as the SOFA score decreased, the number of alarms and the proportions of relevant and relevant plus helpful alarms decreased, and vice versa. Other monitoring devices, including end-tidal $CO_2$ ($EtCO_2$), bladder temperature, and NIBP, were not suited for the univariate analysis because the data size and statistical power were inadequate.

The ART monitor demonstrated a positive correlation between both clinical severity and the proportion of relevant and relevant plus helpful alarms, as well as between clinical severity and the total number of alarms. The $SpO_2$ and ECG monitors demonstrated positive correlations only between the clinical severity and the proportion of relevant and relevant plus helpful alarms.

The inclusion of a regression variable that indicated whether an event occurred during a day or night

shift, in the time-series model, indicated that the time of the alarm did not demonstrate a statistically significant relationship with the SOFA score. In addition, false-negative situation was not recorded during 2352 person-hours of continuous monitoring.

**Discussion**

*GENERAL STATEMENT*

Patients treated in the ICU are surrounded by medical equipment monitoring their vital signs, producing alarms when the measured parameters fall outside of preset "limits." Most types of alarms are not clinically relevant [1-3], lower patient care quality by distracting the medical staff [4-7], and cause patient delirium [9,10]. The present study demonstrated that the proportion of clinically relevant alarms decreased as the patients' status improved, and the total number of alarms can be considerably reduced by evaluating their technical relevance.

Prior to this study, Siebig et al. were the first to record data from a 24-h video monitor, with the help of 2 physicians, to evaluate the clinical relevance of the alarms [13]. This technique reduced the possible bias introduced by bedside evaluation. The same method of evaluation was used in this study, with the added evaluation of the fluctuations in alarm relevance and clinical severity in individual patients. The present study also extends the previous work to a larger cohort.

*ALARM TYPES AND THEIR RELEVANCE*

The vast majority of the alarms triggered in the ICU are either false alarms or are irrelevant for patient treatment. The present study shows that only 6.7% of all alarms triggered in the ICU were actually relevant to patient care, even though 61% were *technically true* alarms. In fact, the relevant alarms were all threshold alarms, as arrhythmia alarms did not yield relevant triggers. These data are similar to the results of multiple prior studies from various institutions, which indicate that approximately 10% of threshold alarms are clinically relevant [1-3,17].

The ART monitor had a positive correlation between a patient's clinical severity and the number and proportion of relevant alarms. In contrast, the $SpO_2$ and ECG monitors only showed positive

correlation between clinical severity and the number of alarms. These findings indicate that the ECG and $SpO_2$ alarms sound regardless of clinical severity. Therefore, ECG and $SpO_2$ alarms are the primary clinically irrelevant alarms, especially with decreasing patient condition severity.

HOW CAN WE REDUCE THE NOISE IN THE ICU?

The present study revealed that ECG and $SpO_2$ monitors were attached to all ICU patients, for safety, starting at ICU admission. Therefore, establishing criteria for removing these devices is difficult. Clinically irrelevant alarms were reduced by 28% by evaluating their technical relevance. Because the ART monitor is often used in the ICU setting, a reduction in the number of clinically irrelevant alarms might be possible by combining the ART waveform with those of the $SpO_2$ monitor and ECG. The number of ART monitor alarms and the proportion of clinically relevant alarms that were associated with patient clinical severity imply that there should be a criterion established to remove this device when the patient's clinical severity has decreased sufficiently.

The most serious problem encountered with these alarms was that while they provided positive predictive values (relevant alarms/all alarms, or relevant plus alerting alarms/all alarms), their sensitivity and specificity could not be obtained. These data could not be obtained because the evaluation of false negatives or true negatives was not possible in cases when the monitor did not set off an alarm in clinical practice. Therefore, manufacturers need to produce alarm devices with high sensitivities in order to avoid any medical accidents. Actually, we did not detect false-negative situations. According to studies by Tsien [3] and Siebig [13] et al., the sensitivity of the current alarms is also close to 100%. However, their specificity, which is important for medical staff, could not be determined.

As a general rule, it is assumed that if the sensitivity and specificity of a given test are constant, positive predictive value (PPV) increases as the (true) prevalence/incidence becomes high. According to this rule, if alarms are triggered constantly, then PPV is higher as the patient severity is higher. Thus, as the patient severity increases, the number of alarms increase, and these alarms include a large number of relevant alarms. In contrast, as the patient severity decreases, the number of alarms

decrease, but these alarms include a small number of relevant alarms. If the significance of medical treatment measured by the alarms is constant, it is desirable that the PPV is constant regardless of the patient's condition. Thus, when the patient severity is low, it is very important to increase PPV strictly according to the standards of sensitivity and specificity.

For many years, many attempts have been made to reduce the number of clinically meaningless alarms, however it remains unresolved. One of the reasons is physicians relatively insensitive to alarm problems, because physicians do not stand by patient bed same as nurses. Thus, physicians, nurses and medical companies need to find a mutually acceptable solution to this matter.

**Conclusion**

Excessive alarms in clinical settings are reportedly linked to lower medical attentiveness and poorer treatment environments. Despite being rarely relevant, these alarms are important. We anticipate the development of an algorithm, which evaluates their technical relevance, thus establishing a criterion that negates the need for such devices.



**Conflicts of interest statement**

The authors declare that they have no conflicts of interest

**Author Contribution**

RI conceived the study. RI and HS designed the analysis plan. RI and HS performed the statistical analyses. RI and HS wrote the first draft of the study. All authors contributed to the design, interpretation of results, and critical revision of the article for intellectually important content.

**Data sharing**

Extra data is available by emailing Ryota Inokuchi .

**Contributorship**

RI conceived the study. RI and HS designed the analysis plan. RI and HS performed the statistical analyses. RI and HS wrote the first draft of the study. All authors contributed to the design, interpretation of results, and critical revision of the article for intellectually important content.

**References**


1. Chambrin MC, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. *Intensive Care Med.* 1999;25(12):1360-1366.
2. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. *Crit Care Med.* 1994;22(6):981-985.
3. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. *Crit Care Med.* 1997;25(4):614-619.
4. Görges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesth Analg.* 2009;108(5):1546-1552.
5. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *Am J Crit Care.* Jan 2010;19(1):28-34
6. Christensen M. Noise levels in a general intensive care unit: a descriptive study. *Nurs Crit Care.* 2007;12(4):188-197.
7. Kam PC, Kam AC, Thompson JF. Noise pollution in the anaesthetic and intensive care environment. *Anaesthesia.* 1994;49(11):982-986.
8. Kahn DM, Cook TE, Carlisle CC, Nelson DL, Kramer NR, Millman RP. Identification and modification of environmental noise in an ICU setting. *Chest.* 1998;114(2):535-540.
9. Zaal IJ, Spruyt CF, Peelen LM, et al. Intensive care unit environment may affect the course of delirium. *Intensive Care Med.* 2012.
10. Radtke FM, Heymann A, Franck M, et al. How to implement monitoring tools for sedation, pain and delirium in the intensive care unit: an experimental cohort study. *Intensive Care Med.* 2012;38(12):1974-1981.
11. Ely EW, Shintani A, Truman B, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *JAMA.* 2004;291(14):1753-1762.
12. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. *Anesth Analg.* 2006;102(5):1525-1537.
13. Siebig S, Kuhls S, Imhoff M, et al. Collection of annotated data in a clinical validation study for alarm algorithms in intensive care--a methodologic framework. *J Crit Care.* 2010;25(1):128-135.
14. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med.* 1985;13(10):818-829.
15. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure

Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707-710.

**16.** Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.

**17.** Koski KJ, Marttila RJ. Transient global amnesia: incidence in an urban population. *Acta Neurol Scand.* 1990;81(4):358-360.

**Figure Legend**

Figure 1. After an evaluation of the technical relevance was made by 2 nurses, an evaluation of clinical relevance was made by 2 intensivists. When evaluating technical relevance, both the threshold and arrhythmia alarms were analyzed; however, the technical aspects of the alarms, such as a connector being unplugged, were not analyzed as these did not require clinical judgment. By combining the movement and procedural data, obtained using the video system, and the data for waveforms of each device, obtained using the monitoring system, the alarms were classified as *technically true* or *technically false*. These alarms were further classified into the following 3 groups by the intensivists: *relevant*, which required an immediate action; *helpful*, which did not require an immediate action; and *not relevant*.

Figure 2. The monitoring devices that triggered alarms the most often were the ART, ECG, and $SpO_2$ monitors.

ART: Direct measurement of arterial pressure, ECG: Electrocardiogram, SpO2: Oxygen saturation, EtCO2: End-tidal carbon dioxide, NIBP: Noninvasive blood pressure

Table 1. Study population characteristics at baseline

Subject description (N = 13)

| | mean ± SD | | |
|---|---|---|---|
| Age (years) | 67.9 ± 16.3 | | |
| Males/females | 9/4 | | |
| | ICU admission | ICU discharge | Model p value |
| APACHE II score | 17.2 ± 8.3 | | |
| SOFA score | 5.92 ± 4.0 | 2.7 ± 1.8 | 0.0147 |
| Total number of monitors | 5.4 ± 0.77 | 4.5 ± 0.78 | 0.0055 |
| Direct measurement of arterial pressure (%) | 76.9 | 15.4 | 0.0008 |
| Electrocardiogram (%) | 100 | 100 | |
| Oxygen saturation monitor (%) | 100 | 100 | |
| End-tidal $CO_2$ ($EtCO_2$) monitor (%) | 61.5 | 38.5 | 0.26 |
| Bladder temperature (%) | 100 | 92.3 | 0.327 |
| Indirect blood pressure measurement (%) | 100 | 100 | |
| Central venous pressure (%) | 0 | 0 | |

Table 2. The total number of all alarms and the number occurring every eight hours.

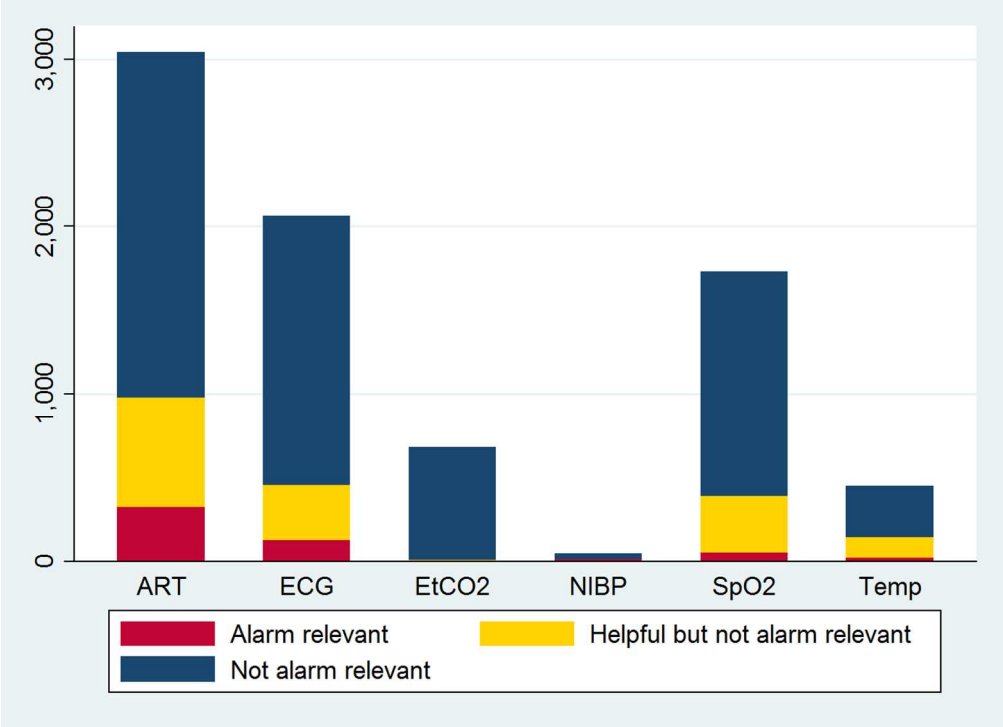
| Alarms (/overall period: 2352 person-hours) | n | Percent of total |
|---|---|---|
| Total numbers | 8013 | |
| Technical annotation | | |
| *Technically true* | 4952 | 61.8% |
| *Technically false* | 2210 | 27.6% |
| Indeterminable | 851 | 10.6% |
| Clinical annotation | | |
| Alarm relevant | 534 | 6.7% |
| Helpful but not alarm relevant | 1457 | 18.2% |
| Not alarm relevant | 6020 | 75.1% |
| Indeterminable | 2 | 0.03% |

| Alarms (/8 h) | |
|---|---|
| Total numbers | 25.4 ± 25.0 |
| Alarm relevant | 1.7 ± 6.9 |
| Helpful but not alarm relevant | 4.6 ± 11.8 |
| Not alarm relevant | 19.1 ± 18.8 |
| Indeterminable | 0.0063 ± 0.11 |

Table 3. Relationship of patients' condition with alarms

| Variable | The total number of alarms | p value | The proportion of relevant alarms (%) | p value | The proportion of relevant and helpful alarms (%) | p value |
|---|---|---|---|---|---|---|
| All monitoring devices | 1.6 ± 0.75 | 0.038 | 2.1 ± 0.45 | < 0.0001 | 3.4 ± 0.83 | < 0.0001 |
| Direct measurement of arterial pressure | 1.6 ± 0.46 | 0.001 | 2.8. ± 0.73 | 0.0002 | 4.2 ± 1.2 | < 0.0001 |
| Electrocardiogram | -0.35 ± 0.42 | 0.412 | 4.4 ± 0.71 | < 0.0001 | 2.5 ± 1.2 | 0.039 |
| Oxygen saturation monitor | 0.35 ± 0.25 | 0.163 | 1.9 ± 0.51 | 0.0003 | 4.9±1.1 | < 0.0001 |

Note: Regression coefficients of severity scores on the (total numbers and proportions of) alarms for all/ each monitoring devices, obtained by the cross-sectional time-series analyses




486x353mm (72 x 72 DPI)

**STROBE STATEMENT checklist of items that should be included in reports of Observational Studies**

| SECTION/TOPIC | Item No. | Checklist Item | Reported on page No. |
|---|---|---|---|
| **TITLE AND ABSTRACT** | 1 | (a) Indicate the study's design with a commonly used term in the title or the abstract | **1, 3** |
| | | (b) Provide in the abstract an informative and balanced summary of what was done and what was found | **3** |
| **INTRODUCTION** | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | **4** |
| Objectives | 3 | State specific objectives, including any pre-specified hypotheses | **4** |
| **METHODS** | | | |
| Study design | 4 | Present key elements of study design early in the paper | **4** |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | **4** |
| Participants | 6 | (a) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>*Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants | **4-5** |
| | | (b) *Cohort study*—For match studies, give matching criteria and number of exposed and unexposed<br>*Case-control study*—For matched studies, give matching criteria and the number of controls per case | |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | **7** |
| Data sources/measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | **5-7** |
| Bias | 9 | Describe any efforts to address potential sources of bias | **13** |
| Study size | 10 | Explain how the study size was arrived at | **9** |

| | | | |
|---|---|---|---|
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | **8** |
| Statistical methods | 12 | (a) Describe all statistical methods, including those used to control for confounding | **8** |
| | | (b) Describe any methods used to examine subgoups and interactions | |
| | | (c) Explain how missing data were addressed | |
| | | (d) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy | |
| | | (e) Describe any sensitivity analyses | |
| **RESULTS** | | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed | **9** |
| | | (b) Give reasons for non-participation at each stage | |
| | | (c) Consider use of a flow diagram | |
| Descriptive data | 14* | (a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders | |
| | | (b) Indicate number of participants with missing data for each variable of interest | |
| | | (c) *Cohort study*—Summarize follow-up time (e.g., average and total amount) | **9** |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time<br>*Case-control study*—Report numbers in each exposure category, or summary measures of exposure<br>*Cross-sectional study*—Report numbers of outcome events or summary measures | **9** |
| Main results | 16 | (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make | **9** |

| | | | |
|---|---|---|---|
| | | clear which confounders were adjusted for and why they were included | |
| | | (b) Report category boundaries when continuous variables were categorized | |
| | | (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | |
| Other analyses | 17 | (d) Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses | **10** |
| **DISCUSSION** | | | |
| Key results | 18 | Summarize key results with reference to study objectives | **11** |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | **13** |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | **12** |
| Generalizability | 21 | Discuss the generalizability (external validity) of the study results | **13** |
| **OTHER INFORMATION** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | **14** |

***Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

# The proportion of clinically relevant alarms decreases as patient clinical severity decreases in intensive care units: A prospective, observational, clinical study

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID: | bmjopen-2013-003354.R1 |
| Article Type: | Research |
| Date Submitted by the Author: | 24-Jul-2013 |
| Complete List of Authors: | Inokuchi, Ryota; The University of Tokyo Hospital, Emergency and Critical Care Medicine<br>Sato, Hajime; National Institute of Public Health, Health Policy and Technology Assessment<br>Nanjo, Yuko; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Echigo, Masahiro; Joint Graduate School of Tokyo Women's Medical University and Waseda University, Cooperative Major in Advanced Biomedical Sciences<br>Tanaka, Aoi; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Ishii, Takeshi; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Matsubara, Takehiro; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Doi, Kent; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Gunshin, Masataka; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Hiruma, Takahiro; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Nakamura, Kensuke; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Shinohara, Kazuaki; Ohta Nishinouchi Hospital, Emergency and Critical care medicine<br>Kitsuta, Yoichi; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Nakajima, Susumu; The University of Tokyo Hospital, Emergency and Critical care medicine<br>Umezu, Mituso; Joint Graduate School of Tokyo Women's Medical University and Waseda University, Cooperative Major in Advanced Biomedical Sciences<br>Yahagi, Naoki; The University of Tokyo Hospital, Emergency and Critical care medicine |
| Primary Subject Heading: | Intensive care |
| Secondary Subject Heading: | Health informatics, Nursing |

| | |
|---|---|
| Keywords: | Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Adult intensive & critical care < ANAESTHETICS, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS |
| | |

**The proportion of clinically relevant alarms decreases as patient clinical severity decreases in intensive care units: A prospective, observational, clinical study**

*(Submitted to BMJ OPEN as an Original Article)*

**Running Title**: Relevant alarms decrease with improved patients status

Ryota Inokuchi MD[1], Hajime Sato MD, MPH, DrPH, PhD[2*], Yuko Nanjo MHS, RN[1], Masahiro Echigo[3] , Aoi Tanaka RN[1], Takeshi Ishii MD[1], Takehiro Matsubara MD, PhD[1], Kent Doi MD, PhD[1], Masataka Gunshin MD[1], Takahiro Hiruma MD[1], Kensuke Nakamura MD[1], Kazuaki Shinohara MD, PhD[4], Yoichi Kitsuta MD, PhD[1], Susumu Nakajima MD, PhD[1], Mitsuo Umezu PhD[3], and Naoki Yahagi MD, PhD[1]

[1] Department of Emergency and Critical Care Medicine, The University of Tokyo Hospital 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

[2] Department of Health Policy and Technology Assessment, National Institute of Public Health 2-3-6 Minami, Wako, Saitama 351-0197, Japan

[3] Cooperative Major in Advanced Biomedical Sciences, Joint Graduate School of Tokyo Women's Medical University and Waseda University, TWIns, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan

[4] Department of Emergency and Critical Care Medicine, Ohta Nishinouchi Hospital, 2-5-20 Nishinouchi, Koriyama, Fukushima 963-8558, Japan

**Acknowledgments**

The authors are deeply grateful to Yugo Tamura for collecting data, and would like to thank Yohei Hashimoto, Kikuo Furuta, and Hiroko Hagiwara for their support. The authors would also like to thank all participating ICU members at The University of Tokyo Hospital for their support.

**Address correspondence to:** Hajime Sato, MD, MPH, DrPH, PhD: Department of Health Policy and Technology Assessment, National Institute of Public Health, 2-3-6 Minami, Wako, Saitama 351-0197, Japan: E-mail: sato.inokuchi2@gmail.com

**Summary**

*Article Focus*

Our objectives were (1) to determine if the proportion and number of clinically relevant alarms differs based on the type of monitoring device, (2) to determine whether patient clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of clinically relevant alarms, and (3) to suggest methods for reducing clinically irrelevant alarms.

*Key Messages*

・The types of devices that alarm the most frequently were those directly measuring arterial pressure (33.5%), oxygen saturation (24.2%), and electrocardiograms (22.9%).

・Clinically relevant alarms decrease proportionally as the clinical severity decreases.

・Clinically irrelevant alarms can be reduced by 21.4% by evaluating technical relevance and wave shapes from the monitoring data, and combining multiple monitoring data.

*Strengths and Limitations*

・We evaluated the technical and clinical relevance of each alarm using 24-h video monitoring. This technique reduced bias introduced by bedside evaluations.

・This study was limited by the small sample size (18 patients, total).

**Grant**

This work was supported by a Grant-in-Aid for Young Scientists (C) (127100000424), and a Health Labour Sciences Research Grant.

**Conflicts of interest statement**

The authors declare that they do not have any conflicts of interest

**Author Contribution**

RI conceived the study. RI and HS designed the analysis plan. RI and HS performed the statistical analyses. RI wrote the first draft of the study. R.I, Y.N., M.E., A.T., T.I., T.M., K.D., M.G., T.H., K.N., Y.K., S.N., and N.Y. contributed to patient management. K.S., M.U., and N.Y. critically reviewed the manuscript. All authors contributed to the design, interpretation of results, and critical revision of the article for intellectually important content.

**Data Sharing Statement**

The technical appendix, statistical code, and dataset are available from the corresponding author at Dryad repository; a permanent, citable, and open access home for the dataset will be provided.

**Abstract**

**Background:** Irrelevant alarms in an intensive care unit (ICU) can harm patient safety. Here, (1) we determined the proportion and number of clinically relevant alarms based on the type of monitoring device, (2) determined whether patient clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of clinically relevant alarms, and (3) suggested methods for reducing clinically irrelevant alarms.

**Methods:** We conducted a prospective, observational, clinical study in a medical ICU at a university hospital. Between January and February 2012, 18 ICU patients were monitored. The alarms, alarm settings, alarm messages, waveforms, and video recordings were acquired in real-time and saved continuously. The alarms were annotated with respect to technical and clinical validity.

**Results:** During 2,697 patient-monitored hours, 11,591 alarms were annotated. Approximately 21.4% of these alarms were classified as technically false. Only 740 (6.4%) alarms were considered clinically relevant. Direct measurement of arterial pressure (33.5%), oxygen saturation (24.2%), and electrocardiogram (22.9%) measurements triggered the most alarms. Positive correlations were established between patient clinical severities and the proportion of relevant alarms. The total number of irrelevant alarms could be reduced by 21.4% by evaluating their technical relevance.

**Conclusions:** We demonstrated that (1) the types of devices that alarm the most frequently were direct measurements of arterial pressure, oxygen saturation, and electrocardiograms, (2) the proportion of clinically relevant alarms decreased as the patients' status improved, and (3) the irrelevance alarms can be considerably reduced by evaluating their technical relevance.

**Keywords:** monitoring device; intensive care unit; alarm algorithm; clinical decision support; patient safety

**Background**

In an intensive care unit (ICU) setting, a large number of medical devices are attached to patients, generating numerous alarm signals every day. Several studies have demonstrated that most of these alarms are not clinically relevant [1-3] and tend to lower the attentiveness of the medical staff, and in turn, lower patient safety [4 5]. In addition, alarm sounds are associated not only with patient delirium [6-10], which increases mortality [11], but also with medical staff memory and judgment disturbances, decreased sensitivity, and exhaustion [6 7]. Many attempts have been made to reduce the number of clinically meaningless alarms by using statistical methods and artificial intelligence systems [12 13]. Some examples include extending the time between the incident and the sounding of the alarm, shutting off alarms prior to performing procedures on patients, and calibrating machines to detect gradual changes in the patient condition. However, alarm devices having high sensitivity and specificity have not been developed because discrepancies remain between the priorities of equipment manufacturers, who are seeking devices with high sensitivity, and those of medical professionals, who desire machines with high specificity.

Previous studies have demonstrated that of the 3 types of alarms—threshold alarms, arrhythmia alarms, and technical alarms—clinical relevance is the lowest for threshold alarms [14]. However, the impact of patient clinical severity on the proportion of clinically relevant alarms remains unknown. Our objectives were (1) to determine if the number and proportion of clinically relevant alarms differs based on the type of monitoring device; (2) to determine whether patient clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of clinically relevant alarms; and (3) to suggest methods for reducing clinically irrelevant alarms. To answer these questions, we used video monitors to collect 24-h continuous data from ICU patients.

**Materials and Methods**

*STUDY SETTING AND PATIENT POPULATION*

This study was conducted in a 6-bed, mixed ICU at the University of Tokyo Hospital, where patients are mainly admitted following ambulance transport. The study ICU is organized in an “I” shape, with 2 individual patient rooms on the west side and 2 double patient rooms on the east side, with a central

monitoring station. The doors to the patient rooms are left open unless procedures are being performed or privacy is required. The unit is staffed with 1 nurse for every 2 patients. Most patients monitored during the study had sepsis, respiratory failure, acute respiratory distress syndrome, multisystem organ failure, renal failure, heart failure, or trauma.

The following inclusion criteria were used to enroll patients in the study: 1) admitted directly to the University of Tokyo Hospital mixed ICU, not stepped-down from other ICUs, and 2) age ≥18 years. Patients were excluded if they were 1) already admitted to this ICU or 2) the patient refused active treatment. This study was approved by the Ethics Committee of the University of Tokyo Hospital, and all patients or their family provided signed informed consent before the beginning of the recordings.

*DATA COLLECTION*

General patient information, such as age, gender, and disease, was recorded. All patients were continuously videotaped using a network of cameras (JVC-Kenwood, V.NET@Web, Tokyo, Japan), attached to the ceiling above each bed, to record patient and/or system manipulations. Each patient was monitored for heart rate, invasive or closely monitored noninvasive arterial blood pressure, respiratory rate, oxygen saturation ($SpO_2$), end-tidal carbon dioxide ($ETCO_2$), and temperature. In addition, any changes in the equipment used for each patient were recorded throughout the study period. In addition, the acute physiology and chronic health evaluation (APACHE II) score [15] was calculated for each patient within 24 h of admission, and the SOFA score [16] was calculated every 8 h. Patient data were pseudonymized, and the electronic files and videos were stored in locked, encrypted, hard drives.

*ALARM SYSTEMS AND SETTINGS*

During the study period, all patients were monitored with a standard cardiovascular monitoring system (BSM-9101 & CNS-9701, Nihon Koden, Tokyo, Japan). The numerical measurements, waveforms, alarms, alarm settings, and alarm messages, were acquired in real-time and saved continuously (CNS-9600 & CAP-2100, Nihon Koden). The alarm information consisted of the parameter causing the alarm and the alarm message (Table 1). The alarm messages were divided into 3

types: threshold alarms, arrhythmia alarms, and technical alarms. The technical alarms indicated technical problems, such as a disconnected probe.

The initial alarm limits and every modification of these during the observation period were registered with corresponding time stamps and automatically recorded (CNS-9600 & CAP-2100, Nihon Koden). Chambrin et al. determined the initial limits for heart rate and systolic arterial pressure by using the rule, "initial value observed during a stable period ± 30%" [1]. This rule was used in this study as well. When prehospital patient heart rates and arterial pressures were not obtained, initial limits were 156/56 mmHg (120/80 ± 30%) for systolic arterial pressure/diastolic pressure and 78 and 43 beats/min (60 ± 30%) for upper and lower heart rate limits, respectively. In addition, the $SpO_2$ limit was 93%, except for patients with chronic obstructive pulmonary disease or acute respiratory distress syndrome, where the limit was 90%; a temperature limit of 38.3°C was also used. After these initial settings, the alarm limits could be modified; any changes were automatically recorded.

*TECHNICAL ANNOTATIONS*

After completion of the data collection for a particular patient, 2 nurses and 2 intensivists, with at least 6 years' experience in intensive care medicine, annotated the data. The 2 nurses first analyzed the technical validity of the alarms, and divided the alarms into 3 categories, *technically true*, *technically false*, and indeterminable. They referred to the multi-monitoring wave shapes or pulse rate when the monitor described alarm messages, not use the video record. Alarms were classified as *technically false*, unnecessary alarms if the monitor referred to other waveforms or pulse rates at the same time. The classifications were defined, in detail, according to the following criteria. For electrocardiogram (ECG), $SpO_2$, direct measurements of arterial pressure, and $ETCO_2$, if the waveform was obviously an artifact produced by movements or procedures, the alarm was determined to be *technically false*. For waveforms in which the origin of the artifact(s) or arrhythmia(s) was uncertain, other waveforms or pulse rates (e.g., a direct measurement of arterial pressure (ART) or $SpO_2$) at the time of alarm generation were also referenced. Alarms that did not meet any of the above criteria were considered *technically true*. All technical evaluations that could not be determined from the relevant monitor's waveform recording were defined as indeterminable. For temperature alarms, all upper and lower

limits of the temperature alarms were defined as t*echnically true*. Finally, for noninvasive blood pressure determinations, if an apparently abnormal value was obtained for the noninvasive blood pressure (NIBP) measurement, the patient's movements and concurrent procedures were also considered. Other values, e.g., ART or $SpO_2$ were also referenced as they may have triggered the upper and lower limit alarms. In such instances, these alarms were considered *technically false*.

*CLINICAL ANNOTATIONS*

After the technical analyses, the 2 physicians divided the alarms into 3 types. These types were relevant alarms, helpful alarms that were not relevant, and irrelevant alarms; these were classified by referring to the video and medical records. In this study, an alarm was defined as relevant when an immediate clinical examination plus diagnostic or therapeutic decision (e.g. ECG, echocardiography, or drug administration) were necessary. When the situation required clinical examination, but did not require a diagnostic or therapeutic decision, it was classified as a helpful alarm but not relevant. Intensivists determining the clinical relevance could see the result of technical validity.

*STATISTICAL ANALYSES*

All included patient characteristics were described using means and SDs for continuous variables, along with medians and ranges. After obtaining the descriptive statistics regarding the alarm counts and their proportions, the bivariate relationship of the alarms (the total number of alarms, and the proportions of relevant alarms) with patient (SOFA) scores were examined by fitting cross-sectional, time-series models for panel data. Alarms from different monitoring devices were examined separately and together. In a preliminary analysis, the numbers and proportions of alarm types were regressed against SOFA scores by fitting either fixed-effects or random-effects models, using the Hausman test. The Hausman test indicated that the random-effects estimates were consistently more appropriate than the fixed-effects estimates [17]. Therefore, the results obtained by the random-effects model were adopted. The interpretation of the statistical significance of relationships was made following multiple comparisons using the Bonferroni method [18]. The NIBP data, were not suited for univariate analysis because the amount of data and statistical power were inadequate.

The intraobserver and interobserver variabilities between the 2 physicians performing the clinical annotations of alarms, and between the 2 nurses performing the technical annotations of the alarms were judged by a Kappa test [19]. To evaluate the intraobserver variability, 300 alarm situations were re-annotated by the same observer after a period of approximately 6 months. Statistical analyses were conducted using STATA Special Edition version 12.1 (StataCorp, College Station, TX, USA).

**Results**

*PATIENT CHARACTERISTICS*

Between January and February 2012, a total of 15,229 alarms were recorded for 20 patients. Two patients were excluded because of their poor clinical condition at the time of admission, and their families' lack of expected benefit from invasive treatment. Therefore, a total of 11,591 alarms for 18 patients were included in this study, corresponding to 2,697 person-monitored hours. The observation time for the cases averaged 150 ± 113 h. Table 2 describes patient characteristics upon admission. During their treatment in the ICU, 66.7% of the patients improved (SOFA scores decreased), while 22.2% deteriorated (SOFA scores increased). The ECG, $SpO_2$, and NIBP devices were attached to all ICU patients, throughout their time in the ICU.

The interobserver variabilities in the technical and clinical annotations, as estimated by the κ coefficient, were 0.98 and 0.68. Similarly, the intraobserver validities were 0.95 and 0.73. These values are within the range of substantial (0.61–0.80) or almost perfect (0.81–1.00) agreement. In addition, false-negative situations were not recorded during the 2,697 patient-monitored hours.

*ALARM CLASSIFICATIONS*

A total of 11,591 alarms were included in the analysis, classified as *technically true* (71%), *technically false* (21.4%), and indeterminable (7.7%) alarms (Fig. 1, Table 3). The overall contribution of each alarm type to the 11,591 alarms is shown in Table 3. Only 6.4% of all alarms were relevant, whereas 32.8% were helpful alarms, but not relevant, and 60.8% of all alarms were irrelevant. During an 8-hour shift, on average, ICU nurses would hear a total of approximately 32 alarms, but only 2 were relevant.

The monitoring devices that triggered alarms the most often were the ART (33.5%), $SpO_2$ (24.2%), and ECG (22.9%) (Fig. 2). The numbers of relevant alarms were 12.4% (ART), 2.4% ($SpO_2$), and 5.3% (ECG).

*EFFECT OF PATIENT STATUS ON THE ALARMS*

The results of the cross-sectional time-series analysis are shown in Table 4. The ART demonstrated a positive correlation between both the SOFA score and the proportion of relevant alarms, as well as between the SOFA score and the total number of alarms, and between the SOFA score and the total number of relevant alarms. The $SpO_2$ and ECG monitors demonstrated positive correlations only between the SOFA score and the proportion of relevant alarms.

All of the devices demonstrated that the SOFA scores had statistically significant positive coefficients when regressed against the total number of relevant alarms ($p < 0.0001$), as well as against the total number of alarms ($p = 0.0061$) and the proportion of relevant alarms ($p < 0.0001$). The results indicated that as the SOFA score decreased, the number of alarms, the number of relevant alarms, and the proportion of relevant alarms decreased; the converse was also true.

The inclusion of a regression variable that indicated whether an event occurred during a day or night shift, in the time-series model, indicated that the time of the alarm did not demonstrate a statistically significant relationship with the SOFA score.

*THE TECHNICAL VALIDITY*

Relevant alarms were comprised of those that were *technically true* and those that were indeterminable, but did not include those that were *technically false*. Thus, the irrelevant alarms could be reduced by 21.4% by evaluating their technical relevance.

**Discussion**

*GENERAL STATEMENT*

ICU patients are surrounded by medical devices that regularly sound alarms, but most of the alarms are not clinically relevant [1-3]. These irrelevant alarms cause a lower quality of patient care by distracting the medical staff [4-7], and contributing to patient delirium [9, 10]. Thus, attempts to reduce the number of clinically irrelevant alarms are important as solutions for this national problem are sought [20]. The present study demonstrated that (1) the devices that alarm the most frequently are ART, $SpO_2$, and ECG; (2) the proportion of relevant alarms decreases as patient status improves; and (3) the irrelevant alarms can be reduced by combining the data for the waveforms or pulse rates of each device.

Prior to this study, Siebig et al. were the first to record data with a 24-h video monitor to, with the help of 2 physicians, evaluate the clinical relevance of alarms [14]. This technique reduced the possible bias introduced by bedside evaluations. The same method of evaluation was used in this study, with the added evaluation of alarm frequency for each device, and the determination of the fluctuations in alarm relevance and clinical severity for individual patients.

*ALARM TYPES AND THEIR RELEVANCE*

The vast majority of the alarms triggered in the ICU are either false alarms or are irrelevant for patient treatment. The present study shows that only 6.4% of all alarms triggered in the ICU were relevant. These data are similar to the results of multiple prior studies from various institutions, which indicated that approximately 10% of alarms are relevant [1-3 21]. The number of alarms that were technically annotated as being indeterminable was 7.7%. When the amplitude of waveforms were small or when arrhythmia indications and noises were mixed, technical annotations were difficult.

The ART alarms had a positive correlation between the SOFA score and the number and proportion of relevant alarms. In contrast, the $SpO_2$ and ECG alarms only showed positive correlations between the SOFA score and the number of alarms. These findings indicate that the $SpO_2$ and ECG alarms sound regardless of clinical severity. Therefore, $SpO_2$ and ECG alarms are the primary, clinically irrelevant alarms, especially in patients with decreasing SOFA scores. However, this study revealed that ECG

and $SpO_2$ devices were attached to all ICU patients, for safety reasons, from the time of their ICU admission. Therefore, establishing criteria for removing these devices would be difficult.

*HOW CAN WE REDUCE THE NOISE IN THE ICU?*

We demonstrated that clinically irrelevant alarms were reduced by 21.4% by evaluating their theoretical technical relevance. When evaluating technical relevance, two nurses combined the data for waveforms or pulse rates for each device. After annotation, their intraobserver and interobserver correlations demonstrated almost perfect agreement and the relevant alarms comprised those that were *technically true* and indeterminable, but not those that were *technically false*. Thus, manufacturers can decrease the number of *technically false* alarms by combining the data from each device. In particular, the ART monitor is often used in the ICU setting, and a reduction in the number of clinically irrelevant alarms might be possible by combining the ART waveform with the data from the $SpO_2$ monitor and ECG.

The number of ART monitor alarms and the proportion of relevant alarms that were associated with the patient SOFA scores implied that there should be a criterion established to remove this device when the SOFA score has decreased to some appropriate level. We found that when the SOFA scores were ≤2, there were no relevant ART alarms. Thus, when SOFA scores are ≤2 and the patient's condition is not likely to change suddenly, the ART device may be removed. As a general rule, if the sensitivity and specificity of a given test are constant, the positive predictive value (PPV) is assumed to increase as the (true) prevalence/incidence becomes higher. According to this rule, if alarms are being triggered constantly, then PPV is higher when the patient illness severity is higher. Thus, as the patient illness severity increases, the number of alarms increase, and these alarms include a large number of relevant alarms. In contrast, as the patient illness severity decreases, the number of alarms decreases, but these alarms include only a small number of relevant alarms. If the significance of medical treatment, measured by the alarms, is constant, the PPV would be more desirably held constant regardless of the patient's condition. Thus, when the patient illness severity is low, an increase in PPV is important, strictly according to the standards of sensitivity and specificity.

*WHY HAS THIS PROBLEM NOT RESOLVED OVER THE PAST DECADE?*

The most serious problem encountered with these alarms was that although they provided PPVs (relevant alarms/all alarms), their sensitivity and specificity cannot be ascertained. These data cannot be ascertained because the evaluation of false negatives and true negatives are not possible in cases where the monitor does not alarm in clinical practice. Therefore, manufacturers need to produce alarmed devices that have higher sensitivities in order to avoid medical accidents. In this study, we did not detect false-negative situations. According to studies by Tsien [3] and Siebig et al. [14], the sensitivity of the current alarms is close to 100%. However, their specificity, which is important for medical staff, could not be determined. Another reason for the failure to reduce the number of clinically irrelevant alarms is that physicians may be relatively insensitive to alarm problems because they do not stand by patient beds as often as nurses. Thus, physicians, nurses, researchers and medical companies need to establish an evidence-based practice model and find a mutually acceptable solution to this matter.

*STUDY LIMITATIONS*

This study has several limitations. The first is that the small sample size was small, only 18 patients. The second limitation is that although a determination could be made regarding whether an alarm was *technically true* or *false*, a strict definition of the clinical annotations was more difficult. There are relevant alarms that require clinical examination, plus diagnostic or therapeutic decision, but this annotation may differ from a definition considered by intensivists. Finally, we did not analyze ventilator and infusion pump alarms, because detailed ventilator alarm messages were not recorded by our system; thus, annotation of their clinical relevance could not be performed. In addition, infusion pump alarms could not connect our system. These irrelevant alarms also need to be decreased [22], and should be the subject of a future study.

**Conclusion**

Excessive alarms in clinical settings are linked to lower medical attentiveness and poorer treatment environments. Manufactures should work to decrease the number of *technically false* alarms by combining waveform data with the device measurement, especially for the ART. Physician should

remove the ART when patient conditions improve sufficiently and they are not likely to change suddenly.

**References**

1. Chambrin MC, Ravaux P, Calvelo-Aros D, et al. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. Intensive Care Med 1999;25(12):1360-6
2. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. Crit Care Med 1994;22(6):981-5
3. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Crit Care Med 1997;25(4):614-9
4. Görges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. Anesth Analg 2009;108(5):1546-52
5. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010;19(1):28-34
6. Christensen M. Noise levels in a general intensive care unit: a descriptive study. Nurs Crit Care 2007;12(4):188-97
7. Kam PC, Kam AC, Thompson JF. Noise pollution in the anaesthetic and intensive care environment. Anaesthesia 1994;49(11):982-6
8. Kahn DM, Cook TE, Carlisle CC, et al. Identification and modification of environmental noise in an ICU setting. Chest 1998;114(2):535-40
9. Zaal IJ, Spruyt CF, Peelen LM, et al. Intensive care unit environment may affect the course of delirium. Intensive Care Med
10. Radtke FM, Heymann A, Franck M, et al. How to implement monitoring tools for sedation, pain and delirium in the intensive care unit: an experimental cohort study. Intensive Care Med 2012;38(12):1974-81
11. Ely EW, Shintani A, Truman B, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. JAMA 2004;291(14):1753-62
12. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. Anesth Analg 2006;102(5):1525-37
13. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010;19(1):28-34
14. Siebig S, Kuhls S, Imhoff M, et al. Collection of annotated data in a clinical validation study for alarm algorithms in intensive care--a methodologic framework. J Crit Care 2010;25(1):128-35
15. Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. Crit Care Med 1985;13(10):818-29

16. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22(7):707-10

17. Greene W. Econom*etric Analysis: Third Edition. Prent*ice Hall, 1997.

18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc 1995;57(1):289-300

19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-74

20. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-77

21. Koski KJ, Marttila RJ. Transient global amnesia: incidence in an urban population. Acta Neurol Scand 1990;81(4):358-60

22. Görges M, Westenskow DR, Markewitz BA. Evaluation of an integrated intensive care unit monitoring display by critical care fellow physicians. J Clin Monit Comput 2012;26(6):429-36

**Figure Legend**

Figure 1. Technical and clinical annotations. After an evaluation of the technical relevance was made by 2 nurses, an evaluation of clinical relevance was made by 2 intensivists.

Figure 2. The numbers and types of different alarms. The monitoring devices that triggered alarms the most often were the ART, ECG, and $SpO_2$ monitors.

ART, Direct measurement of arterial pressure; ECG, Electrocardiogram; $SpO_2$, Oxygen saturation; Temp; Bladder temperature; $ETCO_2$, End-tidal carbon dioxide; NIBP, Noninvasive blood pressure

Table 1. The alarm information consisted of the parameter causing the alarm and the alarm message

| Devices | Threshold alarm | Arrythmia alarm | Technical alarm |
|---|---|---|---|
| **Electrocardiogram (ECG)** | Bradycardia<br>Tachycardia | Asystole<br>ST(II) change<br>Ventricular fibrillation<br>Ventricular tachycardia<br>ventricular premature contraction run | Check electrodes<br>Cannot analyze |
| **Oxygen saturation ($SpO_2$)** | $SpO_2$ | | Not connected<br>Check probe<br>Check probe site<br>Cannnot detect pulse |
| **Direct measurement of arterial pressure (ART)** | ART (Systolic)<br>ART (Diastolic)<br>ART (Mean) | | Not connected<br>Check sensor<br>Check label<br>Cuff occlusion |
| **Noninvasive blood pressure (NIBP)** | NIBP (Systolic)<br>NIBP (Diastolic)<br>NIBP (Mean) | | Not connected<br>Module failure<br>Mead time-out<br>Cannot detect pulse |
| **Capnometer** | $ETCO_2$<br>$CO_2$ (APNEA) | | Not connected<br>Check sensor |
| **Thermometer** | Tblad<br>T2 | | Not connected<br>Check sensor |
| **Central venous pressure monitor** | | | Check sensor |
| **Ventilator** | VENT | | Check sensor |
| **Other** | | | System failure |

$ETCO_2$, end-tidal carbon dioxide; Tblad, bladder temperature

Table 2. Study population baseline characteristics

Subject description (n = 18)

| | Mean ± SD | |
|---|---|---|
| Age | 69.2 ± 14.0 | |
| Male/female | 10/8 (55.6%/44.4%) | |
| | ICU admission | ICU discharge |
| APACHE score | 18.5 ± 8.3 | |
| SOFA score | 6.2 ± 3.8 | 4.1 ± 3.2 |
| The equipment rate of monitoring devices | | |
| Direct measurement of arterial pressure (%) | 77.8 | 33.3 |
| Electrocardiogram (%) | 100 | 100 |
| Oxygen saturation (%) | 100 | 100 |
| End-tidal $CO_2$ ($ETCO_2$) (%) | 61.1 | 44.4 |
| Bladder temperature (%) | 100 | 94.4 |
| Indirect blood pressure measurement (%) | 100 | 100 |

APACHE, acute physiology and chronic health evaluation; SOFA, sequential organ failure assessment

Table 3. The total number of all alarms and the number occurring every eight hours.

| Alarms (/overall period: 2,697 patient-monitored hours) | n | Percent of total |
|---|---|---|
| Total numbers | 11,591 | |
| Technical annotation | | |
| *Technically true* | 8224 | 71.0% |
| *Technically false* | 2479 | 21.4% |
| Indeterminable | 888 | 7.7% |
| Clinical annotation | | |
| Relevant alarm | 740 | 6.4% |
| Helpful, but not relevant, alarm | 3800 | 32.8% |
| Irrelevant | 7049 | 60.8% |
| Indeterminable | 2 | 0.02% |

| Alarms (count/8 h) | Mean ± SD | Median (ranges) |
|---|---|---|
| Total numbers | 31.8 ± 28.6 | 23.5 (1 - 200) |
| Relevant alarm | 2.0 ± 7.7 | 0 (0 - 60) |
| Helpful, but not relevant, alarm | 10.4 ± 13.3 | 6 (0 - 178) |
| Irrelevant | 19.4 ± 20.9 | 13.5 (0 -96) |
| Indeterminable | 0.005 ± 0.1 | 0 (0-2) |

Table 4. Relationship of patient condition with alarm numbers and relevance

| Alarm types | Regression coefficients of severity score (SOFA) [#1 #2] | | | | | |
|---|---|---|---|---|---|---|
| | Total number of alarms | p-value | Total number of relevant alarms | p-value | Percentage of relevant alarms (%) | p-value |
| Direct measurement of arterial pressure | 1.8 ± 0.5 | 0.0001* | 0.6 ± 0.2 | < 0.0001* | 2.2 ± 0.6 | 0.0003* |
| Electrocardiogram | -0.4 ±0.4 | 0.3018 | 0.1 ± 0.1 | 0.066 | 2.4 ± 0.4 | < 0.0001* |
| Oxygen saturation | 0.1 ± 0.3 | 0.7191 | 0.05 ± 0.03 | 0.167 | 0.7 ± 0.2 | 0.0018* |
| Bladder temperature | 0.4 ± 0.2 | 0.0166 | 0.002 ± 0.01 | 0.8704 | -0.1 ±0.4 | 0.7307 |
| End-tidal $CO_2$ | -0.02 ±0.2 | 0.9363 | 0.004 ± 0.004 | 0.4143 | 0.4 ± 0.2 | 0.0726 |

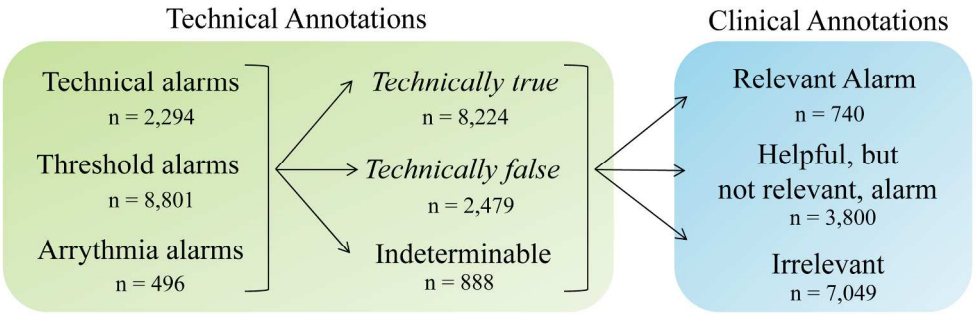
Note:

#1: Only the regression coefficients of severity scores on the (numbers and proportions of) alarms are shown, which were obtained by the cross-sectional time-series analyses (Analysis conducted for each kind of alarms).

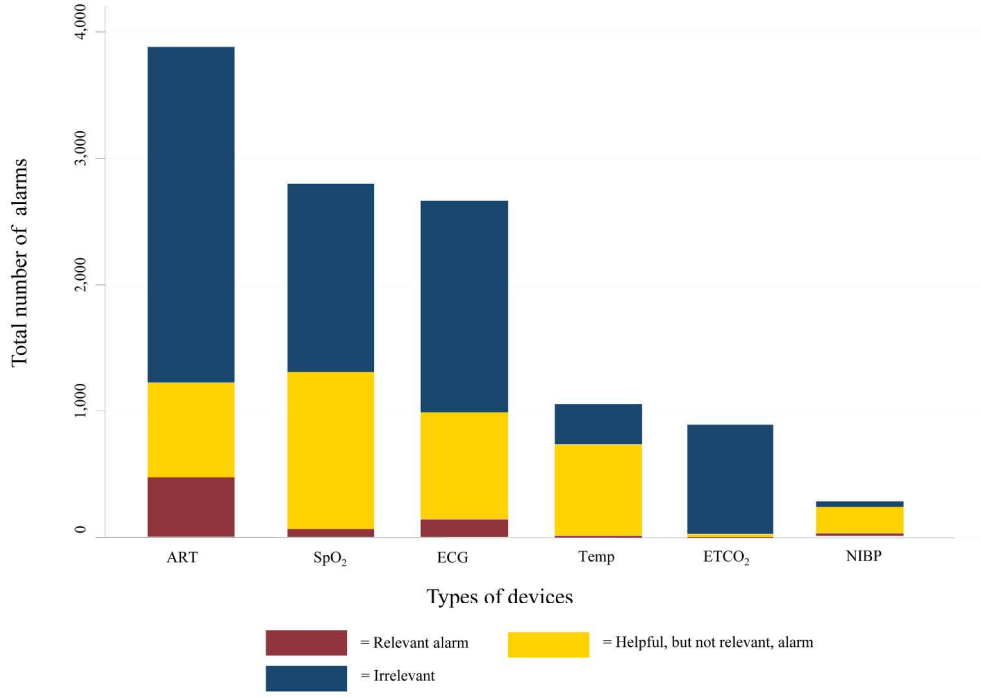
#2: Constant terms were included in the random effect models obtained, but they are not shown.

* Attained statistical significance ($p < 0.05$) after the adjustment for multiple comparisons by Bonferroni method.

SOFA, sequential organ failure assessment




Technical and clinical annotations. After an evaluation of the technical relevance was made by 2 nurses, an evaluation of clinical relevance was made by 2 intensivists.
218x72mm (300 x 300 DPI)




The numbers and types of different alarms. The monitoring devices that triggered alarms the most often were the ART, ECG, and SpO2 monitors.
ART, Direct measurement of arterial pressure; ECG, Electrocardiogram; SpO2, Oxygen saturation; Temp; Bladder temperature; ETCO2, End-tidal carbon dioxide; NIBP, Noninvasive blood pressure
224x163mm (300 x 300 DPI)

**STROBE STATEMENT checklist of items that should be included in reports of Observational Studies**

| SECTION/TOPIC | Item No. | Checklist Item | Reported on page No. |
|---|---|---|---|
| **TITLE AND ABSTRACT** | 1 | (a) Indicate the study's design with a commonly used term in the title or the abstract | **1, 3** |
| | | (b) Provide in the abstract an informative and balanced summary of what was done and what was found | **3** |
| **INTRODUCTION** | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | **4** |
| Objectives | 3 | State specific objectives, including any pre-specified hypotheses | **4** |
| **METHODS** | | | |
| Study design | 4 | Present key elements of study design early in the paper | **4** |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | **4** |
| Participants | 6 | (a) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>*Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants | **4-5** |
| | | (b) *Cohort study*—For match studies, give matching criteria and number of exposed and unexposed<br>*Case-control study*—For matched studies, give matching criteria and the number of controls per case | |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | **7** |
| Data sources/measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | **5-7** |
| Bias | 9 | Describe any efforts to address potential sources of bias | **13** |
| Study size | 10 | Explain how the study size was arrived at | **9** |

| | | | |
|---|---|---|---|
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | **8** |
| Statistical methods | 12 | (a) Describe all statistical methods, including those used to control for confounding | **8** |
| | | (b) Describe any methods used to examine subgoups and interactions | |
| | | (c) Explain how missing data were addressed | |
| | | (d) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy | |
| | | (e) Describe any sensitivity analyses | |
| **RESULTS** | | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed | **9** |
| | | (b) Give reasons for non-participation at each stage | |
| | | (c) Consider use of a flow diagram | |
| Descriptive data | 14* | (a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders | |
| | | (b) Indicate number of participants with missing data for each variable of interest | |
| | | (c) *Cohort study*—Summarize follow-up time (e.g., average and total amount) | **9** |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time<br>*Case-control study*—Report numbers in each exposure category, or summary measures of exposure<br>*Cross-sectional study*—Report numbers of outcome events or summary measures | **9** |
| Main results | 16 | (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make | **9** |

| | | | |
|---|---|---|---|
| | | clear which confounders were adjusted for and why they were included | |
| | | (b) Report category boundaries when continuous variables were categorized | |
| | | (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | |
| Other analyses | 17 | (d) Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses | **10** |
| **DISCUSSION** | | | |
| Key results | 18 | Summarize key results with reference to study objectives | **11** |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | **13** |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | **12** |
| Generalizability | 21 | Discuss the generalizability (external validity) of the study results | **13** |
| **OTHER INFORMATION** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | **14** |

*****Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

**The proportion of clinically relevant alarms decreases as patient clinical severity decreases in intensive care units: A prospective, observational, clinical study**

*(Submitted to BMJ OPEN as an Original Article)*

**Running Title**: Relevant alarms decrease with improved patients status

Ryota Inokuchi MD[1], Hajime Sato MD, MPH, DrPH, PhD[2*], Yuko Nanjo MHS, RN[1], Masahiro Echigo[3] , Aoi Tanaka RN[1], Takeshi Ishii MD[1], Takehiro Matsubara MD, PhD[1], Kent Doi MD, PhD[1], Masataka Gunshin MD[1], Takahiro Hiruma MD[1], Kensuke Nakamura MD[1], Kazuaki Shinohara MD, PhD[4], Yoichi Kitsuta MD, PhD[1], Susumu Nakajima MD, PhD[1], Mitsuo Umezu PhD[3], and Naoki Yahagi MD, PhD[1]

[1] Department of Emergency and Critical Care Medicine, The University of Tokyo Hospital 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

[2] Department of Health Policy and Technology Assessment, National Institute of Public Health 2-3-6 Minami, Wako, Saitama 351-0197, Japan

[3] Cooperative Major in Advanced Biomedical Sciences, Joint Graduate School of Tokyo Women's Medical University and Waseda University, TWIns, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan

[4] Department of Emergency and Critical Care Medicine, Ohta Nishinouchi Hospital, 2-5-20 Nishinouchi, Koriyama, Fukushima 963-8558, Japan

**Acknowledgments**

The authors are deeply grateful to Yugo Tamura for collecting data, and would like to thank Yohei Hashimoto, Kikuo Furuta, and Hiroko Hagiwara for their support. The authors would also like to thank all participating ICU members at The University of Tokyo Hospital for their support.

**Address correspondence to:** Hajime Sato, MD, MPH, DrPH, PhD: Department of Health Policy and Technology Assessment, National Institute of Public Health, 2-3-6 Minami, Wako, Saitama 351-0197, Japan: E-mail: sato.inokuchi2@gmail.com

**Summary**

*Article Focus*

Our objectives were (1) to determine if the proportion and number of clinically relevant alarms differs based on the type of monitoring device, (2) to determine whether patient clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of clinically relevant alarms, and (3) to suggest methods for reducing clinically irrelevant alarms.

*Key Messages*

・The types of devices that alarm the most frequently were those directly measuring arterial pressure (33.5%), oxygen saturation (24.2%), and electrocardiograms (22.9%).

・Clinically relevant alarms decrease proportionally as the clinical severity decreases.

・Clinically irrelevant alarms can be reduced by 21.4% by evaluating technical relevance and wave shapes from the monitoring data, and combining multiple monitoring data.

*Strengths and Limitations*

・We evaluated the technical and clinical relevance of each alarm using 24-h video monitoring. This technique reduced bias introduced by bedside evaluations.

・This study was limited by the small sample size (18 patients, total).

**Grant**

This work was supported by a Grant-in-Aid for Young Scientists (C) (127100000424), and a Health Labour Sciences Research Grant.

**Conflicts of interest statement**

The authors declare that they do not have any conflicts of interest

**Author Contribution**

RI conceived the study. RI and HS designed the analysis plan. RI and HS performed the statistical analyses. RI wrote the first draft of the study. R.I, Y.N., M.E., A.T., T.I., T.M., K.D., M.G., T.H., K.N., Y.K., S.N., and N.Y. contributed to patient management. K.S., M.U., and N.Y. critically reviewed the manuscript. All authors contributed to the design, interpretation of results, and critical revision of the article for intellectually important content.

**Data Sharing Statement**

The technical appendix, statistical code, and dataset are available from the corresponding author at Dryad repository; a permanent, citable, and open access home for the dataset will be provided.

**Abstract**

**Background:** Irrelevant alarms in an intensive care unit (ICU) can harm patient safety. Here, (1) we determined the proportion and number of clinically relevant alarms based on the type of monitoring device, (2) determined whether patient clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of clinically relevant alarms, and (3) suggested methods for reducing clinically irrelevant alarms.

**Methods:** We conducted a prospective, observational, clinical study in a medical ICU at a university hospital. Between January and February 2012, 18 ICU patients were monitored. The alarms, alarm settings, alarm messages, waveforms, and video recordings were acquired in real-time and saved continuously. The alarms were annotated with respect to technical and clinical validity.

**Results:** During 2,697 patient-monitored hours, 11,591 alarms were annotated. Approximately 21.4% of these alarms were classified as technically false. Only 740 (6.4%) alarms were considered clinically relevant. Direct measurement of arterial pressure (33.5%), oxygen saturation (24.2%), and electrocardiogram (22.9%) measurements triggered the most alarms. Positive correlations were established between patient clinical severities and the proportion of relevant alarms. The total number of irrelevant alarms could be reduced by 21.4% by evaluating their technical relevance.

**Conclusions:** We demonstrated that (1) the types of devices that alarm the most frequently were direct measurements of arterial pressure, oxygen saturation, and electrocardiograms, (2) the proportion of clinically relevant alarms decreased as the patients' status improved, and (3) the irrelevance alarms can be considerably reduced by evaluating their technical relevance.

**Keywords:** monitoring device; intensive care unit; alarm algorithm; clinical decision support; patient safety

**Background**

In an intensive care unit (ICU) setting, a large number of medical devices are attached to patients, generating numerous alarm signals every day. Several studies have demonstrated that most of these alarms are not clinically relevant [1-3] and tend to lower the attentiveness of the medical staff, and in turn, lower patient safety [4 5]. In addition, alarm sounds are associated not only with patient delirium [6-10], which increases mortality [11], but also with medical staff memory and judgment disturbances, decreased sensitivity, and exhaustion [6 7]. Many attempts have been made to reduce the number of clinically meaningless alarms by using statistical methods and artificial intelligence systems [12 13]. Some examples include extending the time between the incident and the sounding of the alarm, shutting off alarms prior to performing procedures on patients, and calibrating machines to detect gradual changes in the patient condition. However, alarm devices having high sensitivity and specificity have not been developed because discrepancies remain between the priorities of equipment manufacturers, who are seeking devices with high sensitivity, and those of medical professionals, who desire machines with high specificity.

Previous studies have demonstrated that of the 3 types of alarms—threshold alarms, arrhythmia alarms, and technical alarms—clinical relevance is the lowest for threshold alarms [14]. However, the impact of patient clinical severity on the proportion of clinically relevant alarms remains unknown. Our objectives were (1) to determine if the number and proportion of clinically relevant alarms differs based on the type of monitoring device; (2) to determine whether patient clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of clinically relevant alarms; and (3) to suggest methods for reducing clinically irrelevant alarms. To answer these questions, we used video monitors to collect 24-h continuous data from ICU patients.

**Materials and Methods**

*STUDY SETTING AND PATIENT POPULATION*

This study was conducted in a 6-bed, mixed ICU at the University of Tokyo Hospital, where patients are mainly admitted following ambulance transport. The study ICU is organized in an “I” shape, with 2 individual patient rooms on the west side and 2 double patient rooms on the east side, with a central

monitoring station. The doors to the patient rooms are left open unless procedures are being performed or privacy is required. The unit is staffed with 1 nurse for every 2 patients. Most patients monitored during the study had sepsis, respiratory failure, acute respiratory distress syndrome, multisystem organ failure, renal failure, heart failure, or trauma.

The following inclusion criteria were used to enroll patients in the study: 1) admitted directly to the University of Tokyo Hospital mixed ICU, not stepped-down from other ICUs, and 2) age ≥18 years. Patients were excluded if they were 1) already admitted to this ICU or 2) the patient refused active treatment. This study was approved by the Ethics Committee of the University of Tokyo Hospital, and all patients or their family provided signed informed consent before the beginning of the recordings.

*DATA COLLECTION*

General patient information, such as age, gender, and disease, was recorded. All patients were continuously videotaped using a network of cameras (JVC-Kenwood, V.NET@Web, Tokyo, Japan), attached to the ceiling above each bed, to record patient and/or system manipulations. Each patient was monitored for heart rate, invasive or closely monitored noninvasive arterial blood pressure, respiratory rate, oxygen saturation ($SpO_2$), end-tidal carbon dioxide ($ETCO_2$), and temperature. In addition, any changes in the equipment used for each patient were recorded throughout the study period. In addition, the acute physiology and chronic health evaluation (APACHE II) score [15] was calculated for each patient within 24 h of admission, and the SOFA score [16] was calculated every 8 h. Patient data were pseudonymized, and the electronic files and videos were stored in locked, encrypted, hard drives.

*ALARM SYSTEMS AND SETTINGS*

During the study period, all patients were monitored with a standard cardiovascular monitoring system (BSM-9101 & CNS-9701, Nihon Koden, Tokyo, Japan). The numerical measurements, waveforms, alarms, alarm settings, and alarm messages, were acquired in real-time and saved continuously (CNS-9600 & CAP-2100, Nihon Koden). The alarm information consisted of the parameter causing the alarm and the alarm message (Table 1). The alarm messages were divided into 3

types: threshold alarms, arrhythmia alarms, and technical alarms. The technical alarms indicated technical problems, such as a disconnected probe.

The initial alarm limits and every modification of these during the observation period were registered with corresponding time stamps and automatically recorded (CNS-9600 & CAP-2100, Nihon Koden). Chambrin et al. determined the initial limits for heart rate and systolic arterial pressure by using the rule, "initial value observed during a stable period ± 30%" [1]. This rule was used in this study as well. When prehospital patient heart rates and arterial pressures were not obtained, initial limits were 156/56 mmHg (120/80 ± 30%) for systolic arterial pressure/diastolic pressure and 78 and 43 beats/min (60 ± 30%) for upper and lower heart rate limits, respectively. In addition, the $SpO_2$ limit was 93%, except for patients with chronic obstructive pulmonary disease or acute respiratory distress syndrome, where the limit was 90%; a temperature limit of 38.3°C was also used. After these initial settings, the alarm limits could be modified; any changes were automatically recorded.

*TECHNICAL ANNOTATIONS*

After completion of the data collection for a particular patient, 2 nurses and 2 intensivists, with at least 6 years' experience in intensive care medicine, annotated the data. The 2 nurses first analyzed the technical validity of the alarms, and divided the alarms into 3 categories, *technically true*, *technically false*, and indeterminable. They referred to the multi-monitoring wave shapes or pulse rate when the monitor described alarm messages, not use the video record. Alarms were classified as *technically false*, unnecessary alarms if the monitor referred to other waveforms or pulse rates at the same time. The classifications were defined, in detail, according to the following criteria. For electrocardiogram (ECG), $SpO_2$, direct measurements of arterial pressure, and $ETCO_2$, if the waveform was obviously an artifact produced by movements or procedures, the alarm was determined to be *technically false*. For waveforms in which the origin of the artifact(s) or arrhythmia(s) was uncertain, other waveforms or pulse rates (e.g., a direct measurement of arterial pressure (ART) or $SpO_2$) at the time of alarm generation were also referenced. Alarms that did not meet any of the above criteria were considered *technically true*. All technical evaluations that could not be determined from the relevant monitor's waveform recording were defined as indeterminable. For temperature alarms, all upper and lower

limits of the temperature alarms were defined as t*echnically true*. Finally, for noninvasive blood pressure determinations, if an apparently abnormal value was obtained for the noninvasive blood pressure (NIBP) measurement, the patient's movements and concurrent procedures were also considered. Other values, e.g., ART or $SpO_2$ were also referenced as they may have triggered the upper and lower limit alarms. In such instances, these alarms were considered *technically false*.

*CLINICAL ANNOTATIONS*

After the technical analyses, the 2 physicians divided the alarms into 3 types. These types were relevant alarms, helpful alarms that were not relevant, and irrelevant alarms; these were classified by referring to the video and medical records. In this study, an alarm was defined as relevant when an immediate clinical examination plus diagnostic or therapeutic decision (e.g. ECG, echocardiography, or drug administration) were necessary. When the situation required clinical examination, but did not require a diagnostic or therapeutic decision, it was classified as a helpful alarm but not relevant. Intensivists determining the clinical relevance could see the result of technical validity.

*STATISTICAL ANALYSES*

All included patient characteristics were described using means and SDs for continuous variables, along with medians and ranges. After obtaining the descriptive statistics regarding the alarm counts and their proportions, the bivariate relationship of the alarms (the total number of alarms, and the proportions of relevant alarms) with patient (SOFA) scores were examined by fitting cross-sectional, time-series models for panel data. Alarms from different monitoring devices were examined separately and together. In a preliminary analysis, the numbers and proportions of alarm types were regressed against SOFA scores by fitting either fixed-effects or random-effects models, using the Hausman test. The Hausman test indicated that the random-effects estimates were consistently more appropriate than the fixed-effects estimates [17]. Therefore, the results obtained by the random-effects model were adopted. The interpretation of the statistical significance of relationships was made following multiple comparisons using the Bonferroni method [18]. The NIBP data, were not suited for univariate analysis because the amount of data and statistical power were inadequate.

The intraobserver and interobserver variabilities between the 2 physicians performing the clinical annotations of alarms, and between the 2 nurses performing the technical annotations of the alarms were judged by a Kappa test [19]. To evaluate the intraobserver variability, 300 alarm situations were re-annotated by the same observer after a period of approximately 6 months. Statistical analyses were conducted using STATA Special Edition version 12.1 (StataCorp, College Station, TX, USA).

**Results**

*PATIENT CHARACTERISTICS*

Between January and February 2012, a total of 15,229 alarms were recorded for 20 patients. Two patients were excluded because of their poor clinical condition at the time of admission, and their families' lack of expected benefit from invasive treatment. Therefore, a total of 11,591 alarms for 18 patients were included in this study, corresponding to 2,697 person-monitored hours. The observation time for the cases averaged 150 ± 113 h. Table 2 describes patient characteristics upon admission. During their treatment in the ICU, 66.7% of the patients improved (SOFA scores decreased), while 22.2% deteriorated (SOFA scores increased). The ECG, $SpO_2$, and NIBP devices were attached to all ICU patients, throughout their time in the ICU.

The interobserver variabilities in the technical and clinical annotations, as estimated by the κ coefficient, were 0.98 and 0.68. Similarly, the intraobserver validities were 0.95 and 0.73. These values are within the range of substantial (0.61–0.80) or almost perfect (0.81–1.00) agreement. In addition, false-negative situations were not recorded during the 2,697 patient-monitored hours.

*ALARM CLASSIFICATIONS*

A total of 11,591 alarms were included in the analysis, classified as *technically true* (71%), *technically false* (21.4%), and indeterminable (7.7%) alarms (Fig. 1, Table 3). The overall contribution of each alarm type to the 11,591 alarms is shown in Table 3. Only 6.4% of all alarms were relevant, whereas 32.8% were helpful alarms, but not relevant, and 60.8% of all alarms were irrelevant. During an 8-hour shift, on average, ICU nurses would hear a total of approximately 32 alarms, but only 2 were relevant.

The monitoring devices that triggered alarms the most often were the ART (33.5%), $SpO_2$ (24.2%), and ECG (22.9%) (Fig. 2). The numbers of relevant alarms were 12.4% (ART), 2.4% ($SpO_2$), and 5.3% (ECG).

*EFFECT OF PATIENT STATUS ON THE ALARMS*

The results of the cross-sectional time-series analysis are shown in Table 4. The ART demonstrated a positive correlation between both the SOFA score and the proportion of relevant alarms, as well as between the SOFA score and the total number of alarms, and between the SOFA score and the total number of relevant alarms. The $SpO_2$ and ECG monitors demonstrated positive correlations only between the SOFA score and the proportion of relevant alarms.

All of the devices demonstrated that the SOFA scores had statistically significant positive coefficients when regressed against the total number of relevant alarms ($p < 0.0001$), as well as against the total number of alarms ($p = 0.0061$) and the proportion of relevant alarms ($p < 0.0001$). The results indicated that as the SOFA score decreased, the number of alarms, the number of relevant alarms, and the proportion of relevant alarms decreased; the converse was also true.

The inclusion of a regression variable that indicated whether an event occurred during a day or night shift, in the time-series model, indicated that the time of the alarm did not demonstrate a statistically significant relationship with the SOFA score.

*THE TECHNICAL VALIDITY*

Relevant alarms were comprised of those that were *technically true* and those that were indeterminable, but did not include those that were *technically false*. Thus, the irrelevant alarms could be reduced by 21.4% by evaluating their technical relevance.

**Discussion**

*GENERAL STATEMENT*

ICU patients are surrounded by medical devices that regularly sound alarms, but most of the alarms are not clinically relevant [1-3]. These irrelevant alarms cause a lower quality of patient care by distracting the medical staff [4-7], and contributing to patient delirium [9, 10]. Thus, attempts to reduce the number of clinically irrelevant alarms are important as solutions for this national problem are sought [20]. The present study demonstrated that (1) the devices that alarm the most frequently are ART, $SpO_2$, and ECG; (2) the proportion of relevant alarms decreases as patient status improves; and (3) the irrelevant alarms can be reduced by combining the data for the waveforms or pulse rates of each device.

Prior to this study, Siebig et al. were the first to record data with a 24-h video monitor to, with the help of 2 physicians, evaluate the clinical relevance of alarms [14]. This technique reduced the possible bias introduced by bedside evaluations. The same method of evaluation was used in this study, with the added evaluation of alarm frequency for each device, and the determination of the fluctuations in alarm relevance and clinical severity for individual patients.

*ALARM TYPES AND THEIR RELEVANCE*

The vast majority of the alarms triggered in the ICU are either false alarms or are irrelevant for patient treatment. The present study shows that only 6.4% of all alarms triggered in the ICU were relevant. These data are similar to the results of multiple prior studies from various institutions, which indicated that approximately 10% of alarms are relevant [1-3 21]. The number of alarms that were technically annotated as being indeterminable was 7.7%. When the amplitude of waveforms were small or when arrhythmia indications and noises were mixed, technical annotations were difficult.

The ART alarms had a positive correlation between the SOFA score and the number and proportion of relevant alarms. In contrast, the $SpO_2$ and ECG alarms only showed positive correlations between the SOFA score and the number of alarms. These findings indicate that the $SpO_2$ and ECG alarms sound regardless of clinical severity. Therefore, $SpO_2$ and ECG alarms are the primary, clinically irrelevant alarms, especially in patients with decreasing SOFA scores. However, this study revealed that ECG

and $SpO_2$ devices were attached to all ICU patients, for safety reasons, from the time of their ICU admission. Therefore, establishing criteria for removing these devices would be difficult.

*HOW CAN WE REDUCE THE NOISE IN THE ICU?*

We demonstrated that clinically irrelevant alarms were reduced by 21.4% by evaluating their theoretical technical relevance. When evaluating technical relevance, two nurses combined the data for waveforms or pulse rates for each device. After annotation, their intraobserver and interobserver correlations demonstrated almost perfect agreement and the relevant alarms comprised those that were *technically true* and indeterminable, but not those that were *technically false*. Thus, manufacturers can decrease the number of *technically false* alarms by combining the data from each device. In particular, the ART monitor is often used in the ICU setting, and a reduction in the number of clinically irrelevant alarms might be possible by combining the ART waveform with the data from the $SpO_2$ monitor and ECG.

The number of ART monitor alarms and the proportion of relevant alarms that were associated with the patient SOFA scores implied that there should be a criterion established to remove this device when the SOFA score has decreased to some appropriate level. We found that when the SOFA scores were ≤2, there were no relevant ART alarms. Thus, when SOFA scores are ≤2 and the patient's condition is not likely to change suddenly, the ART device may be removed.

As a general rule, if the sensitivity and specificity of a given test are constant, the positive predictive value (PPV) is assumed to increase as the (true) prevalence/incidence becomes higher. According to this rule, if alarms are being triggered constantly, then PPV is higher when the patient illness severity is higher. Thus, as the patient illness severity increases, the number of alarms increase, and these alarms include a large number of relevant alarms. In contrast, as the patient illness severity decreases, the number of alarms decreases, but these alarms include only a small number of relevant alarms. If the significance of medical treatment, measured by the alarms, is constant, the PPV would be more desirably held constant regardless of the patient's condition. Thus, when the patient illness severity is low, an increase in PPV is important, strictly according to the standards of sensitivity and specificity.

*WHY HAS THIS PROBLEM NOT RESOLVED OVER THE PAST DECADE?*

The most serious problem encountered with these alarms was that although they provided PPVs (relevant alarms/all alarms), their sensitivity and specificity cannot be ascertained. These data cannot be ascertained because the evaluation of false negatives and true negatives are not possible in cases where the monitor does not alarm in clinical practice. Therefore, manufacturers need to produce alarmed devices that have higher sensitivities in order to avoid medical accidents. In this study, we did not detect false-negative situations. According to studies by Tsien [3] and Siebig et al. [14], the sensitivity of the current alarms is close to 100%. However, their specificity, which is important for medical staff, could not be determined. Another reason for the failure to reduce the number of clinically irrelevant alarms is that physicians may be relatively insensitive to alarm problems because they do not stand by patient beds as often as nurses. Thus, physicians, nurses, researchers and medical companies need to establish an evidence-based practice model and find a mutually acceptable solution to this matter.

*STUDY LIMITATIONS*

This study has several limitations. The first is that the small sample size was small, only 18 patients. The second limitation is that although a determination could be made regarding whether an alarm was *technically true* or *false*, a strict definition of the clinical annotations was more difficult. There are relevant alarms that require clinical examination, plus diagnostic or therapeutic decision, but this annotation may differ from a definition considered by intensivists. Finally, we did not analyze ventilator and infusion pump alarms, because detailed ventilator alarm messages were not recorded by our system; thus, annotation of their clinical relevance could not be performed. In addition, infusion pump alarms could not connect our system. These irrelevant alarms also need to be decreased [22], and should be the subject of a future study.

**Conclusion**

Excessive alarms in clinical settings are linked to lower medical attentiveness and poorer treatment environments. Manufactures should work to decrease the number of *technically false* alarms by

combining waveform data with the device measurement, especially for the ART. Physician should remove the ART when patient conditions improve sufficiently and they are not likely to change suddenly.

**References**

1. Chambrin MC, Ravaux P, Calvelo-Aros D, et al. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. Intensive Care Med 1999;25(12):1360-6
2. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. Crit Care Med 1994;22(6):981-5
3. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Crit Care Med 1997;25(4):614-9
4. Görges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. Anesth Analg 2009;108(5):1546-52
5. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010;19(1):28-34
6. Christensen M. Noise levels in a general intensive care unit: a descriptive study. Nurs Crit Care 2007;12(4):188-97
7. Kam PC, Kam AC, Thompson JF. Noise pollution in the anaesthetic and intensive care environment. Anaesthesia 1994;49(11):982-6
8. Kahn DM, Cook TE, Carlisle CC, et al. Identification and modification of environmental noise in an ICU setting. Chest 1998;114(2):535-40
9. Zaal IJ, Spruyt CF, Peelen LM, et al. Intensive care unit environment may affect the course of delirium. Intensive Care Med
10. Radtke FM, Heymann A, Franck M, et al. How to implement monitoring tools for sedation, pain and delirium in the intensive care unit: an experimental cohort study. Intensive Care Med 2012;38(12):1974-81
11. Ely EW, Shintani A, Truman B, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. JAMA 2004;291(14):1753-62
12. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. Anesth Analg 2006;102(5):1525-37
13. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010;19(1):28-34
14. Siebig S, Kuhls S, Imhoff M, et al. Collection of annotated data in a clinical validation study for alarm algorithms in intensive care--a methodologic framework. J Crit Care 2010;25(1):128-35
15. Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. Crit Care Med 1985;13(10):818-29

16. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22(7):707-10

17. Greene W. Econom*etric Analysis: Third Edition. Prent*ice Hall, 1997.

18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc 1995;57(1):289-300

19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-74

20. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-77

21. Koski KJ, Marttila RJ. Transient global amnesia: incidence in an urban population. Acta Neurol Scand 1990;81(4):358-60

22. Görges M, Westenskow DR, Markewitz BA. Evaluation of an integrated intensive care unit monitoring display by critical care fellow physicians. J Clin Monit Comput 2012;26(6):429-36

**Figure Legend**

Figure 1. Technical and clinical annotations. After an evaluation of the technical relevance was made by 2 nurses, an evaluation of clinical relevance was made by 2 intensivists.

Figure 2. The numbers and types of different alarms. The monitoring devices that triggered alarms the most often were the ART, ECG, and $SpO_2$ monitors.

ART, Direct measurement of arterial pressure; ECG, Electrocardiogram; $SpO_2$, Oxygen saturation; Temp; Bladder temperature; $ETCO_2$, End-tidal carbon dioxide; NIBP, Noninvasive blood pressure

Table 1. The alarm information consisted of the parameter causing the alarm and the alarm message

| Devices | Threshold alarm | Arrythmia alarm | Technical alarm |
|---|---|---|---|
| **Electrocardiogram (ECG)** | Bradycardia<br>Tachycardia | Asystole<br>ST(II) change<br>Ventricular fibrillation<br>Ventricular tachycardia<br>ventricular premature contraction run | Check electrodes<br>Cannot analyze |
| **Oxygen saturation ($SpO_2$)** | $SpO_2$ | | Not connected<br>Check probe<br>Check probe site<br>Cannnot detect pulse |
| **Direct measurement of arterial pressure (ART)** | ART (Systolic)<br>ART (Diastolic)<br>ART (Mean) | | Not connected<br>Check sensor<br>Check label<br>Cuff occlusion |
| **Noninvasive blood pressure (NIBP)** | NIBP (Systolic)<br>NIBP (Diastolic)<br>NIBP (Mean) | | Not connected<br>Module failure<br>Mead time-out<br>Cannot detect pulse |
| **Capnometer** | $ETCO_2$<br>$CO_2$ (APNEA) | | Not connected<br>Check sensor |
| **Thermometer** | Tblad<br>T2 | | Not connected<br>Check sensor |
| **Central venous pressure monitor** | | | Check sensor |
| **Ventilator** | VENT | | Check sensor |
| **Other** | | | System failure |

$ETCO_2$, end-tidal carbon dioxide; Tblad, bladder temperature

Table 2. Study population baseline characteristics

Subject description (n = 18)

| | Mean ± SD | |
|---|---|---|
| Age | 69.2 ± 14.0 | |
| Male/female | 10/8 (55.6%/44.4%) | |
| | ICU admission | ICU discharge |
| APACHE score | 18.5 ± 8.3 | |
| SOFA score | 6.2 ± 3.8 | 4.1 ± 3.2 |
| The equipment rate of monitoring devices | | |
| Direct measurement of arterial pressure (%) | 77.8 | 33.3 |
| Electrocardiogram (%) | 100 | 100 |
| Oxygen saturation (%) | 100 | 100 |
| End-tidal $CO_2$ ($ETCO_2$) (%) | 61.1 | 44.4 |
| Bladder temperature (%) | 100 | 94.4 |
| Indirect blood pressure measurement (%) | 100 | 100 |

APACHE, acute physiology and chronic health evaluation; SOFA, sequential organ failure assessment

Table 3. The total number of all alarms and the number occurring every eight hours.

| Alarms (/overall period: 2,697 patient-monitored hours) | n | Percent of total |
|---|---|---|
| Total numbers | 11,591 | |
| Technical annotation | | |
| *Technically true* | 8224 | 71.0% |
| *Technically false* | 2479 | 21.4% |
| Indeterminable | 888 | 7.7% |
| Clinical annotation | | |
| Relevant alarm | 740 | 6.4% |
| Helpful, but not relevant, alarm | 3800 | 32.8% |
| Irrelevant | 7049 | 60.8% |
| Indeterminable | 2 | 0.02% |

| Alarms (count/8 h) | Mean ± SD | Median (ranges) |
|---|---|---|
| Total numbers | 31.8 ± 28.6 | 23.5 (1 - 200) |
| Relevant alarm | 2.0 ± 7.7 | 0 (0 - 60) |
| Helpful, but not relevant, alarm | 10.4 ± 13.3 | 6 (0 - 178) |
| Irrelevant | 19.4 ± 20.9 | 13.5 (0 -96) |
| Indeterminable | 0.005 ± 0.1 | 0 (0-2) |

Table 4. Relationship of patient condition with alarm numbers and relevance

| Alarm types | Regression coefficients of severity score (SOFA) [#1 #2] | | | | | |
|---|---|---|---|---|---|---|
| | Total number of alarms | p-value | Total number of relevant alarms | p-value | Percentage of relevant alarms (%) | p-value |
| Direct measurement of arterial pressure | 1.8 ± 0.5 | 0.0001* | 0.6 ± 0.2 | < 0.0001* | 2.2 ± 0.6 | 0.0003* |
| Electrocardiogram | -0.4 ±0.4 | 0.3018 | 0.1 ± 0.1 | 0.066 | 2.4 ± 0.4 | < 0.0001* |
| Oxygen saturation | 0.1 ± 0.3 | 0.7191 | 0.05 ± 0.03 | 0.167 | 0.7 ± 0.2 | 0.0018* |
| Bladder temperature | 0.4 ± 0.2 | 0.0166 | 0.002 ± 0.01 | 0.8704 | -0.1 ±0.4 | 0.7307 |
| End-tidal $CO_2$ | -0.02 ±0.2 | 0.9363 | 0.004 ± 0.004 | 0.4143 | 0.4 ± 0.2 | 0.0726 |

Note:

#1: Only the regression coefficients of severity scores on the (numbers and proportions of) alarms are shown, which were obtained by the cross-sectional time-series analyses (Analysis conducted for each kind of alarms).

#2: Constant terms were included in the random effect models obtained, but they are not shown.

* Attained statistical significance ($p < 0.05$) after the adjustment for multiple comparisons by Bonferroni method.

SOFA, sequential organ failure assessment